Proceedings of the IATUL Conferences 1970 IATUL Proceedings

The Future of Information Provision

D. J. Hillman
Lehigh University

THE FUTURE OF INFORMATION PROVISION*

by

DONALD J. HILLMAN

1. Libraries and Procognitive Systems

The role of the library vis-a-vis science and technology is different in function and scope from that of the library serving the arts and humanities, although recognition of the differing roles took some time to evolve. Thus the first librarians of the Library attached to the famous Museum of Alexandria were all men of letters, and it was not until the third century B.C. that the classification and investigation of scientific books was entrusted to the care of a man of science, viz., Eratosthenes of Cyrene.

The circumstances surrounding the appointment and tenure of Eratosthenes were not always of the best. Although he was a superb geographer who estimated the earth's circumference within an accuracy of 1%, and a talented mathematician whose famous "sieve" provided an algorithm for enumerating the prime numbers, Eratosthenes was nevertheless given two nicknames, "Beta" and"Pentathlos". "Beta" means number two, or second rate, while "Pentathlos" means someone skilled at the five games, and derivatively "jack of all trades". The eclecticism of Eratosthenes is what made his administration of the library so inventive and apt for the science and technology of his time. Nowadays, we would call his library "interdisciplinary" and "service-oriented", words which would have made the more pure-minded of his users shudder, although I shall have occasion to employ these words frequently in what I have to say later. The main point I wish to make is that Eratosthenes was probably the first scientific and technological librarian whose understanding of the genuine transdisciplinary needs of science far exceeded that of his users. His two nicknames show that the scientists and scholars of the third century B.C. were already very jealous of one another and were all too ready to deflate those whose superiority they misunderstood.

Although the modern counterparts of Eratosthenes run less risk of abuse from their users, or at least experience abuse in more subtle forms, it is nevertheless true that much of what must be done to keep the technological library responsive to changing needs will transform its role in as dramatic a fashion as brought about by Eratosthenes in converting the Library at Alexandria from a literature-oriented to a science-oriented institution. The library of the future will bear little resemblance to present-day libraries, largely because books and their physical accommodation will be lower on the list of priorities than they are now. Many of the problems to be encountered are, of course, already present in one form or another. These can be conveniently grouped under the heading of "library automation" and concern such problems as catalogue construction, acquisitions

* Grateful acknowledgement is made to the National Science Foundation for Grant GN-845 on which this paper is based, and for Grant GN-869 to enable the author to attend the Conference.

activities, serial control, and circulation. As Kilgour (1) has recently observed:

> "Library computerisation has begun with computerisation of existing procedures. Indeed, it is difficult to see how the course could have been otherwise; technology advances from an existing base."

However, the kind of library I envisage will be very far removed from the traditional conception, with its conventional activities of acquisition and circulation on either side of catalogue construction. To be sure, these activities will continue, but in a radically different form. Where the library of the future will least resemble the library of the present is in its handling of books, and it is this development that poses the greatest problem of all.

Part of the problem, curiously enough, results from linguistic and cultural entrenchment. The English word "library" has, after all, a very straightforward etymology, through the Latin "librarius" from "liber", book. It is a truism that the term "library" connotes books to virtually every English-speaking person, so that to talk of libraries without books is to invite misunderstanding of one's intentions or doubts as to one's sanity. Yet it is precisely the notion of a library without books that I wish to promulgate at this conference, and I hope in the remainder of this paper to justify the concept both for future implementation and also in terms of a library/information system currently under development at Lehigh University.

Before commenting explicitly on this development, it is instructive to consider the likely societal and sociological effects of designing and operating bookless libraries. To begin with, it will be immensely difficult to persuade society that the phrase "library without books" represents anything at all, except possibly a majestic but forbidding building, resembling a courthouse or other official building, usually erected (in the United States at least) through the philanthropy of Andrew Carnegie, and now as a result of some fiscal catastrophe without books of any kind on its shelves. In other words, people would invariably think of a library without books as a completely conventional library that happened to have no books, whereas what I am referring to is a nonconventional library designed to house no books. A predictable communication breakdown will occur at this point, largely because the word "library" is so deeply entrenched in our language and our conceptual scheme as a book-storing and disseminating institution. Yet concepts change as our understanding of them changes, and it will need nothing less than a drastic change in our understanding of the interface between technology and culturally-entrenched institutions like the library to bring about the developments I have in mind.

The path towards this new understanding is a hazardous one, with not the least of our difficulties arising in the form of widespread refusal to recognise the library of the future as any kind of library at all. I mentioned earlier the linguistically irresistible association between libraries and books, and this is a powerful force indeed. The history of science is full of episodes in which fresh understanding and new insights were blocked, in some cases for centuries, by purely linguistic prejudices. Consider, for example, the evolution of our concept of space. Aristotle, and nearly two thousand years later, Descartes, both denied the possibility of a vacuum. Descartes held that the essence of matter is extension, and that the relation of extension to material body is, in scholastic language, that of attribute to substance. Since an attribute can exist only as the attribute of some substance, it follows that there can be no such thing as extension without matter, or, in other words, there cannot be a vacuum. Thus space, by a purely linguistic argument, becomes a so-called plenum.

This view persisted until it was overthrown by the discoveries of Newton at the end of the seventeenth century. The Cartesian view held its ground a little longer in France, it should be added, to the extent that Voltaire, returning from Newton's funeral in 1727, said that he had left space a vacuum in London and found it a plenum in Paris. But even in Paris the English view eventually prevailed.

The Newtonian concept was that space is infinite, uniform, continuous, immovable, and Euclidean in its geometry. Leibniz, as might be expected, criticised the Newtonian doctrine, arguing that space is an order of coexistences. The excellence of this definition was subsequently realised by a countryman of Leibniz's in formulating his special theory of relativity.

The Newtonian view of space as absolute nothingness prevailed until the nineteenth century, when physicists began to feel uncomfortable about finding a medium for supporting the wave-motion of light. They argued that every predicate must have a subject, and that motion cannot take place without anything being moved. In perhaps the most striking example of linguistic prejudice ever in the history of science, the physicists, in the memorable words of Lord Balfour, invented the ether in order to provide a nominative for the verb "to undulate". Space, by a linguistic argument, became a plenum again, there to remain until Einstein joined space with time into a 4-dimensional continuum and put a halt to further unbridled linguistic speculation.

Before this had happened, Immanuel Kant, in the century following Newton, had further confused the issue by propounding a theory of space in which all geometrical propositions can be deduced a priori from intuition with absolute certainty. For much of the nineteenth century, Kant's view lent credence to the belief that Euclidean geometry was infallibly true, thus preventing Gauss from publishing his discovery of non-Euclidean geometry because he feared the "outcry of the blockheads". This magnificent discovery was practically ignored for forty years, and the disservice rendered by Kant to mathematics almost, but not quite, matches his numerous malfeasances in philosophy.

Enough, I think, has now been said about the forces of linguistic entrenchment to show how difficult it is to overcome the barriers of language and culture in gaining acceptance for conceptual modifications.

It is entirely likely that something like the acrimonious debate over the notion of space will be paralleled by a similar discussion over the concept of the future library, and it will be a long time, if ever, before people get used to the idea of a library without books.

Yet such a development is absolutely necessary for the health and vitality of the future library. Let me quote from Licklider's excellent monograph Libraries of the Future in which he says:

> "As a medium for the display of information, the printed page is superb. It affords enough resolution to meet the eye's demand. It presents enough information to occupy the reader for a convenient quantum of time. It offers great flexibility of font and format. It lets the reader control the mode and rate of inspection. It is small, light, movable, cuttable, clippable, pastable, replicable, disposable, and inexpensive. Those positive attributes all relate, as indicated, to the display function. The tallies that could be made for the storage, organisation, and retrieval functions are less favourable.

"When printed pages are bound together to make books or journals, many of the display features of the individual pages are diminished or destroyed. Books are bulky and heavy. They contain much more information than the reader can apprehend at any given moment, and the excess often hides the part he wants to see. Books are too expensive for universal private ownership, and they circulate too slowly to permit the development of an efficient public utility. Thus, except for use in consecutive reading - which is not the modal application in the domain of our study - books are not very good display devices. In fulfilling the storage function, they are only fair. With respect to retrievability they are poor. And when it comes to organising the body of knowledge, or even to indexing and abstracting it, books by themselves make no active contribution at all (2)."

It is Licklider's thesis that any view of a library encompassing the idea of books on shelves is bound to run into trouble, attributable partly to logistic difficulties but more importantly to difficulties stemming from the passiveness of the printed page. He argues:

"We need to substitute for the book a device that will make it easy to transmit information without transporting material, and that will not only present information to people but also process it for them, following procedures they specify, apply, monitor, and if necessary revise and reapply. To provide those services, a meld of library and computer is evidently required (3)."

Computerisation, as noted earlier, has so far been applied to the traditional tasks of cataloguing, acquisition, and circulation, but the fundamental problem of the future library will not be solved by merely improving library organisation at these levels. What is needed is an entire restructuring of the library and its function relative to the dynamic corpus of knowledge. Licklider has coined the term "procognitive system" to designate a system facilitating man's interaction with transformable information, and it is as a first step towards a procognitive system that I wish to describe the LEADERMART Project now under development at Lehigh University.

In keeping with Licklider's blueprint, the LEADERMART system rejects the printed page as a long-term storage device, and excludes physical books as passive repositories of information. All information handled by LEADERMART meets Licklider's criterion of transformability, i.e., rephrasability without significant loss, thus excluding works of art and literature. With this by way of introduction, let me now turn to a discussion of LEADERMART as it is now being implemented at Lehigh.

2. The LEADERMART System

A. Objectives

The LEADERMART* System at Lehigh University is a campus-centred information system whose development is being sponsored by the National Science Foundation under Grant GN-845. It is designed to provide an operating service for a network of on-campus scientific and engineering research centres. These centres were

* LEADER is an acronym for Lehigh Answer to Demand for Efficient Retrieval: MART refers to the Mart Library. - Ed.

established to promote inter-disciplinary research in eight areas, and to encourage interaction among the science and engineering disciplines. LEADERMART is thus science-oriented rather than discipline-oriented, so that data items of widely different origin may be communicated across disciplinary boundaries.

The task of developing the retrieval system has been assigned to the Center for Information Science, while the responsibility for operating the system will be assumed by the Mart Library of Science and Engineering.

The immediate objectives of the proposed development are:

(i) to provide a science-oriented information service for the interdisciplinary research centers at Lehigh University;

(ii) to furnish the centers with an on-line capacity to interact with several different data bases;

(iii) to make the information system an efficient switching mechanism for the transdisciplinary use of research information;

(iv) to provide a variety of retrieval procedures for users;

(v) to determine the importance of various cost effectiveness and utility parameters of the system for the purpose of operational optimisation.

The long range objectives of the development are:

(i) to establish the foundations of a science-oriented information system serving a larger network of users;

(ii) to train research scientists and engineers in the use of adaptive systems without formal classroom instruction;

(iii) to modernise and enhance the service function of a new science and engineering library;

(iv) to stimulate computer assisted research in science and engineering.

B. <u>System Design</u>

The LEADERMART system is a man-machine symbiosis featuring multi-level interactions between an information processing facility and Lehigh research centre personnel. Operating within the organisational framework of the Mart Library of Science and Engineering, the system has four major subsystems controlled by a supervisor. These subsystems govern the flow of, and perform transformations on, the data being transmitted through system channels. The supervisor controls data transmission through the four subsystems, and schedules user/system interactions. Since the information circulating through the system is organised and structured in a definite sequence, the four subsystems govern the flow, generation, and transformation of primary, secondary, tertiary, and quaternary information, respectively.

i. Primary Information Subsystem

Primary information consists of journal articles, technical reports, serials, test results and other data. This information is produced by many sources external to the information system, and is acquired on the basis of a selection procedure. Each user group specifies the nature of the primary information pertinent to its research needs, and a corresponding data base is defined. These several data bases are largely discipline-oriented.

Primary information flows into the Mart Library as a result of its serials and other document acquisition. A well-defined portion of this information also flows into the LEADERMART information system through the user-defined filter at the source information/primary information subsystem interface (See Figure 1).

ii. Secondary Information Subsystem

The primary information in the first subsystem is now directed by the supervisor to the secondary information subsystem, where it is analysed, organised, classified, and stored in files in such a way as to make it available to users whose information needs dictate crossings of various disciplinary boundaries, e.g. physics and metallurgy for information regarding the preparation and properties of materials for solid-state devices.

Secondary information is here defined as the result of various transformations of primary information whereby the latter is (a) organised into concept-oriented rather than discipline-oriented schemes, and (b) compacted or selectively rearranged so as to become responsive to transdisciplinary needs. The forms and structures developed within the secondary information subsystem, such as indexes and other bibliographic data, abstracts, excerpts, synopses, are typical vehicles of secondary information. In addition, the conceptual schemes established for the primary information, resulting in several different file structures, are crucially necessary for communication across disciplinary boundaries. In this respect, the proposed information system is globally-oriented in its secondary information subsystem, whose processing is a completely internal activity. That is, there is no interface between the secondary information subsystem and the external environment.

iii. Tertiary Information Subsystem

The function of this subsystem is to select information judged to be pertinent to specific problems or missions and to deliver such information to appropriate user groups. The number of different ways in which information can be selected and delivered is subject to the joint control of the user and the supervisor, and various retrieval responses, including document and reference retrieval, data retrieval, SDI, etc., are available.

Tertiary information is defined as the result of a controlled dialogue between the user and the system involving secondary information as processed, transformed, repackaged, compressed or expanded, etc., during an on-line, real-time negotiation process or an off-line SDI or demand procedure. The tertiary information subsystem has an extensive interface with users.

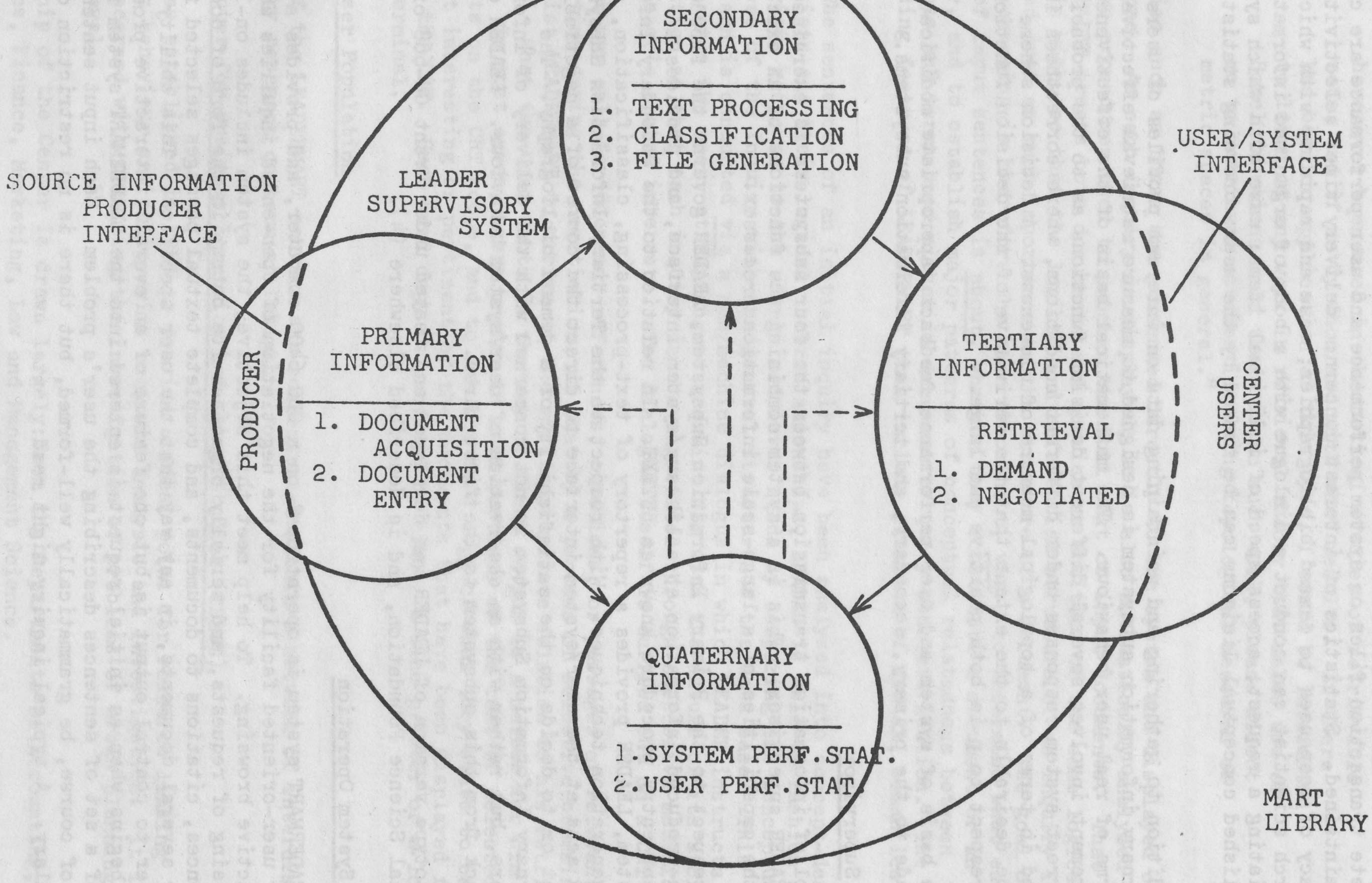

FIGURE 1. SUBSYSTEM STRUCTURE

iv. Quaternary Information Subsystem

Complete transaction files on system performance and user performance are created and maintained. Statistics of interest concern: delivery time, selectivity, and accuracy of responses to demand bibliographies; ease and rapidity with which a research scientist can conduct a dialogue with a body of organised information in negotiating a request; access speed of individual items; ease with which system-established conceptual liaisons can be traced by the user; browsing statistics, etc.

In addition to gathering and maintaining data on interest profiles of users, the quaternary information subsystem is designed to measure retrieval effectiveness in terms of real user behaviour. The mathematical basis of the effectiveness measurement involves several different decision functions as to the probability p of correct system responses under different conditions, where correctness is defined in terms of a topological measure of closeness. A decision scheme will then be desirable to the extent that the derivative of the decision function with respect to p is both positive and large.

On the basis of system and user performance feedback, appropriate modifications are made to the primary, secondary, and tertiary information subsystems.

v. Supervisor

Control of information transmission between the four subsystems is exercised by the LEADER supervisor. This is a system combining the functions of an executive with the capabilities of a large-scale information processor.

With respect to the Primary Information Subsystem, LEADER governs the flow of source-produced information at a library/system interface, and provides a document entry procedure known as LETEXT. In relation to the Secondary Information Subsystem, LEADER provides a repertory of text-processing, classification, and file generation techniques. With respect to the Tertiary Information Subsystem, LEADER acts at the user/system interface to direct the course of a negotiable request or to decide on the satisfiability of a demand bibliography. The Quaternary Information Subsystem is not concerned with the delivery of information to users, but rather with an observation of user/system behaviour. LEADER controls feedback from this subsystem to the first three.

A prototype version of LEADER was developed and tested under Grant GN-668 of the National Science Foundation, and is described elsewhere (4).

C. System Operation

The LEADERMART system is operational on a CDC 6400 computer, and provides a very highly user-oriented facility for the negotiation of open-ended inquiries and interactive browsing. To help meet this objective, the system includes on-line processing of requests, and serially organises its output in the form of document references, citations to documents, and complete textual passages selected from one or several documents, in any way that the user specifies. This ability of the user to control output is but one feature of an overall interactive procedure which begins when an initial request is entered into the LEADERMART system in the form of a set of sentences describing the user's problem. Each input sentence must, of course, be grammatically well-formed, but there is no restriction on vocabulary. A typical inquiry might read:

"I would like to know whether modular bounded functionals have ever been used in theoretical studies of retrievable sets, and if so by whom and with what results. If there has been no application of this type, I would be interested to learn of any work in retrieval theory that makes use of Borel functions. If there is no such work, please direct me to retrieval studies involving topological measures or metric spaces in general."

Inquiries such as these are presented directly to the system and displayed on a CRT scope. As each inquiry is displayed, it is also automatically analysed by the same procedures used to process the full text of input documents. That is to say, LEADER treats both documents and queries as entities of the same logical type to begin with, so that the logical and referential structure of an inquiry is accorded just as much importance as the structure of a document. The goal of text processing is therefore the same throughout, viz., to determine what each group of input sentences is about, whether they constitute a document or an inquiry, and to establish major patterns of conceptual relatedness between documents and terms used either in document or query characterisation. The text-processing features of LEADERMART thus include elements of syntax, semantics, and logic.

After the sentences of an initial inquiry have been analysed into concept-denoting expressions and their logical interrelationships, LEADER is able to fashion an appropriate response to the user's retrieval needs by comparing the conceptual structure of the inquiry with the general structure of the data base. This comparison is conducted via a man/machine dialogue in which LEADER instructs and interrogates the user, attempting to acquaint him with the nature of its stored information so that each inquiry can be negotiated through successive modifications of the user's stated interests. The dialogue itself is carried out on a CRT scope.

The user may call for document references, citations, or passages of relevant text at any time during the negotiation, so that by a process of selective browsing he may assist LEADERMART to arrive at the most appropriate solution to his retrieval problem. What the user wants in the way of final output is a matter for him to decide. In most cases, users prefer to read portions of the text of selected documents on the CRT scope, and to ask for hard copy of what they state to be the most interesting or pertinent of the documents that have been displayed for them on the CRT. Such hard copy is provided by either a line printer or a low-speed terminal.

D. User Population

Users of the LEADERMART service are concentrated within a network of interdisciplinary research centers at Lehigh.

i. Center for Business Economics and Urban Studies (CBEUS)

CBEUS was established in 1965 to provide a focal point for research linking business behaviour and policy with economic analysis and investigation. It is also concerned with promoting the investigation of problems of urban planning and renewal.

Membership of the Center is drawn largely from the Departments of Accounting, Economics, Finance, Marketing, Law and Management Science.

ii. <u>Center for Surface and Coatings Research (CSCR)</u>

CSCR was established in 1966 to promote research into solid/gas and solid/liquid interactions. The solid/gas effort is concerned largely with corrosion, particularly the mechanism whereby metals corrode, lose strength, and fracture under stress. The solid/liquid studies are aimed principally at chemical coatings, particularly colloidal dispersions of pigments in polymeric fluids.

The Departments of Chemistry, Chemical Engineering, and Mechanics are principally involved in the research activities of CSCR.

iii. <u>Computing Center</u>

The Computing Center was formed in 1957 as the University Computing Laboratory. It has a CDC 6400, and provides computer service for the solution of instructional and research problems. The center also serves as a laboratory for departmental courses and research in computer theory, programming, and information systems.

iv. <u>Fritz Engineering Laboratory</u>

Founded in 1909, the Fritz Engineering Laboratory is the largest Lehigh centre for research in structures, structural mechanics, materials, hydraulics and fluid mechanics, structural model analysis, soil mechanics, and sanitation. Research projects are sponsored by national research councils, industrial corporations and associations, private companies, and state and federal government agencies.

The Laboratory is associated primarily with the Department of Civil Engineering, and houses the world's largest universal hydraulic testing machine capable of applying a 5,000,000 lb load to tension or compression members up to 40 feet in length and flexure specimens up to 120 feet long.

Fritz Laboratory participates in a world-wide exchange of research information, and maintains a special library of technical papers appropriate to its several research fields.

v. <u>Center for Information Science (CIS)</u>

CIS was established in 1962 for R and D projects in information systems and their operation. In its system development role, CIS provides the project staff for the LEADERMART service. In its dual role of user, CIS maintains its collection of current literature in information science, which now numbers some 6,000 items.

vi. <u>Center for Marine and Environmental Studies (CMES)</u>

CMES was established in 1962 to foster interdisciplinary studies of ocean, physical, chemical, biological, and geological processes. The Center is actively concerned with research in marine microbiology and marine bacteria, and is particularly interested in the biological effects of thermal pollution.

The members of CMES are drawn from the Departments of Geological Science, Biology, Mechanics, Chemistry, and Chemical Engineering.

vii. Materials Research Center (MRC)

MRC was established in 1962 to encourage interaction among the science and engineering disciplines with an interest in materials. The Center has five laboratories, viz., Advanced Materials Laboratory, Mechanical Behaviour Laboratory, Polymer Laboratory, Physical Ceramics Laboratory, and Electron Microscopy Laboratory In addition, there are several associated laboratories physically located within departments affiliated with MRC for: Engineering Structure Analysis, Surface Chemistry, Stress Corrosion, Magnetic Materials, Crystal Growing and Zone Processing Manufacturing Processes, Hydrothermal Synthesis, Solid State Investigations, and Environmental Science Studies.

The members of MRC are drawn from the Department of Metallurgy and Materials Science Chemistry, Physics, Chemical Engineering, Civil Engineering, Electrical Engineering, Geological Sciences, Mechanics and Mechanical Engineering, and Industrial Engineerin

The Materials Liaison Program of MRC serves as a means of exchange of research information between members of MRC and their industrial and government counterparts.

viii. Center for the Application of Mathematics (CAM)

CAM was established in 1965 to foster interdisciplinary research related to the application of mathematics, to draw on other disciplines for pertinent mathematical problems, and to promote nonconventional uses of computers.

Since mathematics is fundamental to all scientific and engineering activities on campus, the members of CAM represent a complete cross-section of mathematical interests at Lehigh. Departments most actively involved include: Chemical Engineering, Civil Engineering, Electrical Engineering, Industrial Engineering, Management Science, Mathematics, Mechanical Engineering and Mechanics, Philosophy, Information Science, Physics, and Psychology.

E. Operational Responsibility

The LEADERMART system is being developed for subsequent operation and administration by the Mart Library of Science and Engineering.

The Lehigh University library system consists of two major buildings housing approximately 500,000 volumes. The Linderman Library incorporates the original building and a more modern structure erected in 1929. Annual accessions number about 13,000 volumes, and more than 4,000 current periodicals and serials are received. The library is a depository for a wide selection of U.S. government documents, and has particularly strong collections in the physical and natural sciences, mathematics, and engineering. The Honeyman Collection of rare books is distinguished for its strength in the history of science.

The Mart Library of Science and Engineering was completed in March 1969, has space for 151,000 volumes, and currently houses 85,000 volumes in the fields of engineering, mathematics, and the natural and physical sciences, allowing the present Linderman Library space to increase its holdings in the humanities and social sciences. The Mart Library includes facilities for an all-night study room, automated circulation system, programmed learning room, computer and teletype consoles, closed-circuit television, and other instrumentation.

The nature and strength of the cooperation between the Center for Information Science and the University Libraries make it highly appropriate to develop the described LEADERMART system as part of the service mission of the Mart Library to the scientific and engineering community at Lehigh. Not only is CIS historically linked with the Lehigh Libraries, but both organisations share the common goal of making a data base of scientific and engineering information maximally accessible to faculty and students.

F. Interactions with Procognitive Systems

I have argued in this paper that neolibrary systems of the future must be designed to expedite man's interaction with bodies of transformable information within the store of scientific and technological knowledge. All such systems will be computer-based, and will need to separate the information in documents from the printed page. The inter-active servicing of information requests takes the form of a dialogue between man and a procognitive system, the intent of which is to arrive at a joint decision as to the nature, volume, pertinence, and utility of the information to be formatted, retrieved, displayed, and put to work.

Although the procedures involved in negotiating a request are many and varied, they are sufficiently similar in logical structure to permit of a single unified view of decision operations. What is common to all decision procedures is a principle whereby items of one logical type are compared with items of another logical type. Thus, topic-denoting items occuring in an inquiry may be compared with topic-denoting items used in document processing. What makes one kind of decision procedure different from another is a change in the form such a comparison may take, ranging from simple matching of terms, on the one hand, to a detailed investigation of the logical and conceptual environments of different terms, on the other. We may also provide a comparative ordering of decision procedures along a scale of directness. If we regard a request negotiation as a continuum, then user promptings and system responses may be thought of as connected events within the continuum arranged in order of increasing directness of retrieval. For example, retrieval of bibliographic references, such as author, title and journal, is a very indirect way of servicing requests, in that the user is told that he can probably find the information he wants in those documents mentioned by the retrieved references. Such indirectness of access to stored information is in fact a necessary consequence of document or reference retrieval, whose epistemological requirements are limited to knowledge that something has been said about a given topic in certain documents. It is for this reason that reference retrieval events occur early in LEADERMART dialogues, with later events providing even more direct access to the required information. As the dialogue becomes less and less exploratory, so the form of retrieval approaches knowledge concerning what has been said about a previously identified topic or set of topics. "Knowing that" and "knowing what" are therefore at opposite ends of the retrieval continuum, although not all requests need to reach the "knowing what" stage.

From this description of request processing in LEADERMART, it is clear that decision procedures must provide the basis for a dialogue in which the user and the system can trade suggestions during the negotiation of an input inquiry up to the time the user calls for a halt. In the course of the dialogue, the system will furnish different kinds of retrieval output corresponding to different stages of negotiation, ranging from exploratory probings of user interest to specific details of subject matter. Although no two dialogues are identical in the way they are conducted, each is nevertheless built out of the same logical components, so that a unified formal basis for the set of decision procedures may be constructed. As mentioned earlier, the fundamental components of a decision process are comparisons between items of the same or different logical types. Differences between actual retrieval negotiations may therefore be explained as differences between types and orderings of comparisons and their corresponding outputs. The most elementary type of comparison requires simple matching of terms, while the most advanced and in many respects most interesting type of comparison involves consideration of the conceptual connections among topics as established by the system *vis-a-vis* user-supplied connections. In its highest manifestation, information retrieval deals with comparisons between different conceptual schemes, and the LEADERMART system has been designed to project describable portions of users' conceptual schemes onto the conceptual scheme of an information source. The decision component of LEADERMART thus provides a kind of road map enabling users to trace important conceptual paths among topics either globally (i.e. system-established), or by local overlays (i.e. user-supplied connections). I take these features to be fundamental in the future of information provision.

REFERENCES

(1) Kilgour, Frederick G. "Library Automation", Ch.10, Annual Review of Information Science and Technology, Vol 4, 1969, p328.

(2) Licklider, J.C.R. Libraries of the Future, The M.I.T. Press, 1965, p4-5.

(3) Licklider, J.C.R. Ibid., p6.

(4) Hillman, Donald J. and Kasarda, Andrew J. "The LEADER Retrieval System", 1969 Spring Joint Computer Conference, AFIPS Conf. Proc. Vol 34, 1969, p447-455.

DISCUSSION

J.D. MACK: Would Prof. Hillman tell us how errors in linguistic analysis or processing are corrected.

D.J. HILLMAN: We have developed a subroutine to correct spelling errors. Some problems do arise in the preparation of documents because they are written in incorrect English.

C.G. WOOD: May one assume that English is the only source language used?

D.J. HILLMAN: So far, we have spent five years on the program for linguistic analysis and it has been limited to the English language. (In reply to an unidentified questioner Prof. Hillman added that it was sometimes easier to analyse long sentences than short ones. - Ed)

C.C. PARKER: Can Prof. Hillman give an idea of the time involved in processing the full text input and in searching?

D.J. HILLMAN: For an average 1000 sentence article the analysis would take about 8 minutes but the subsequent internal process is very fast. With regard to searching a user may perhaps spend 45 minutes at the terminal while negotiating with the computer, but again the actual processing time is very small, probably only about 45 seconds.

R.A. WALL: How are such large matrix data bases handled in what, presumably, are necessarily sequential stores.

DISCUSSION CONT'D

D.J. HILLMAN: Sequential stores are not used but direct address to records on disks. Processing time is minimal because matrix inversion is not practised; instead, a series of related (and weighted) terms is extracted and constitutes a 'unique probability vector of an ergodic Markov Chain' for terms related to the enquiry. There is one matrix per genus. The genera themselves are arrived at by syntactic analysis.

J.E. DAVIES: Would Prof. Hillman expand a little on the description of the information collected in the quaternary stage? How is performance of both user and system assessed and how are the references weighted for suitability?

D.J. HILLMAN: With regard to evaluation of system performance we keep statistics on how well the syntactic analysis has been conducted, and on how well, in a 'pure systems' sense, operations have been carried out. User performance is really a measure of how well the user has adapted to the system. During the interrogation of the computer the conversation is interrupted and terms are suggested. The overlap of computer terms with the user's terms gives a measure of the user's success in conversing with the machine. A complete transaction file will show how well there has been adaptation to the system. The weighting of terms for relevance is done by the computer. Relevance is judged by the strength of the links with other terms and the operational definition of relevance is based on this. User relevance can be judged by the use he makes of the documents supplied but, of course, only applies in the context of an individual user.

R.A. WALL: Prof. Hillman has said that the processing time per sentence for syntactic analysis is now down to half a second. May we have a little more detail about the present method of initial input of data? How long will it be before the product of typesetting can be used for input, directly or by optical character recognition.

D.J. HILLMAN: There are now three possible methods of input:

(a) taped abstracts - various available tape services are used;
(b) full text - the on-line LETEXT system permits visual editing via a display console;
(c) Photo-offset, i.e. the computer typesetting tapes, can also be used as input.

Optical character recognition possibilities, however, have not been investigated because of the expense of the equipment.

R.A. WALL: In the early experiments, reported at the Cranfield Conference several years ago, it was necessary to exclude general documents to avoid, in effect, 'short-circuiting' the matrices. Would Professor Hillman tell us if this is still the case, or are there now some safeguards in this system?

D.J. HILLMAN: In a sense, selection is still done, but the full text documents input is user defined and many are, in fact, survey or review articles. Amendment of the matrices can be delayed until use determines the probable effects. Documents which are not used can be dropped from the system.

A.G. MYATT: Is it possible for Prof. Hillman to estimate the future growth rate of the total disk store?

D.J. HILLMAN: By the end of 1972 we estimate there will be 500×10^6 characters in fast access disk storage, this being half of the total capacity of the computer. Older material on tape will go into 'archive' storage of infinite capacity.

J. LUBANS: What is the projected fixed cost of operating LEADERMART when the Library takes it over in 1972? Would Prof. Hillman care to forecast when we shall see the 'procognitive system' (or non-book library) actually in being.

D.J. HILLMAN: The annual cost is estimated to be $250,000, some of which we hope to meet with funds from local industry and research groups. Funded research at the University will also contribute something. The development of a procognitive system is bound to take place in phases. At present LEADERMART is not concerned with books, and cannot become interested until all books are photocomposed. Perhaps Licklider's suggestion of 2000 A.D. is a fair guess.

F. TAFT: Can Prof. Hillman say how they are dealing with the copyright question?

D.J. HILLMAN: We are ignoring those problems at present.

J.D. MACK: It would be interesting to know the number of staff needed to run the LEADERMART project?

D.J. HILLMAN: The project staff consists of 6 senior members and 12 systems analysts and programmers with more of the latter during the summer. The average length of programming experience is 7 years/person and these people are able to write extremely sophisticated programs.